09/555032

422 Rec'd PCT/PTO 22 MAY 2000

INS A1

4/prts

Method and Device for Voice Recognition

INS A2

A

The present invention relates to a method and a device for voice recognition.

INS A3

A voice recognition system is taught in ~~[1]~~ A4. This also contains a basic introduction to the components included in the voice recognition system, as well as important techniques that are common in voice recognition.

In a known voice recognition system, a degree of accuracy – that is, a measure of a quality of the recognition – is predetermined. The user must now make do with this system, even when a lower degree of accuracy would suffice for his application, though he would achieve a higher operating speed in the bargain.

The principle of pruning a search space is known from ~~[2]~~ A5 (see chapter 3.3.3, page 40). This is a matter of "trimming" the search space, or rather a method for reducing a number of search paths of the search space, whereby the least promising search paths are cut off. To this end, first a search path with minimal costs (optimal search path) is established. Then, all search paths (branches of the search tree) whose costs are above the minimum inclusive of an added prescribed evaluation quantity, which is referred to as the pruning threshold, are cut off. For a detailed explanation of the pruning: ~~[2]~~ A6, p. 40ff; particularly Figure 16 on page 41.

When a pruning threshold is used, it is not known how many search paths will remain in the search tree. If one wishes to maintain the number of these remaining search paths at a predetermined level, the pruning threshold is dynamically adapted.

A histogram pruning is taught in ~~[3] and [4]~~ A7. Here, a predetermined number of "best" search paths are used – that is, search paths with a high probability of occurrence – in that frequencies of the search paths are evaluated in the form of a histogram. The pruning threshold is dynamically modified.

An acoustic look-ahead in the search tree (term of art: fast look ahead) is taught in ~~[5] and [6].~~ A8

A

The idea pursued in the acoustic look-ahead (also referred to as fast preselection) is based on the characteristic of a language that all words are composed of a limited inventory of linguistic subunits (e.g. phonemes, half-syllables). An acoustic weighting is now performed for these linguistic subunits "in advance". Only those combinations ~~[sic]~~ of linguistic subunits are tracked whose acoustic weights are below a predetermined threshold. An advantage in the weighting outlay is that, for a low number of linguistic subunits, a measure of the agreement of a speech signal that is to be recognized, on one hand, and a target quantity, on the other hand, is computed in advance and used as a basis for deciding whether a large part of the search tree should be excluded from consideration. Simply put, this means that more search paths in the search tree are reserved than are added in by the prediction. A profit like this grows larger the higher the ratio of new word beginnings to a number of linguistic subunits becomes. This ratio increases with the number of linguistic subunits, or respectively, words, that are to be recognized (lexicon size).

An advantage of the method of acoustic look-ahead consists in the regularity of the algorithms for computing the corresponding scores. Since there are no branchings in the search space due to word ends, syntactic nodes, and so on, the schema of the calculation of the scores is regular. Exactly for this reason, it is possible to use this type of method for an implementation in hardware also.

The prediction of the scores A9 (term of art: fast match scores) is possible because the actual search lags behind the current extracted scores of the speech signal by a fixed number of time windows. Using the current scores, the prediction of the scores of further linguistic subunits is performed (see A10 ~~[5]~~, p. 65, Figure 33).

This type of look-ahead is also carried out in language models (see A11 ~~[6]~~).

The principle of language model look-ahead is to take the probabilities existing in the language model into account in the search process as soon as possible, and also in the associated pruning. This is achieved by a factorization of the probabilities in the language model. A detailed description with a formal notation is contained in ~~[6]~~. A12

A

Finally, a threshold for selecting computing ~~[sic]~~ distance parameters is taught in ~~[7]~~. A13 Such selection methods are generally multi-stepped. First, a rough calculation is performed using a part of the distances. In the next step, those distances are determined which are close to the best distance of the first computational step with respect to an interval score. This interval score can be varied via a threshold, whereby the computing outlay for determining the distance parameters is varied.

INS A14

~~The object of the invention is to create a~~ method and a device for adjusting the accuracy of the voice recognition system. A15

This object is achieved by ~~the features of the independent claims.~~ A16

A

A method for adjusting the accuracy of a voice recognition system is set forth, in which the accuracy is determined by a prescribable input quantity. With the aid of this input quantity, values for system parameters of the voice recognition system are computed, preferably by a computer. With the aid of these values, the voice recognition system is adjusted. This is preferably accomplished by the computer automatically.

The advantage of the invention is that it makes the accuracy of the voice recognition system adaptable and at the same time adjustable to laypersons. Depending on the application, or respectively, on the computing power available for the voice recognition system on the computer, various requirements can be placed on the quality of the voice recognition system simply by adapting the input quantity.

One development consists in calculating the values for the system parameters of the voice recognition system from the input quantity in accordance with a mapping specification. This mapping specification can be converted using a table.

By setting the input quantity, it is also possible to automatically compute the values of the system parameters of the voice recognition system that are linked to these and thus to automatically make the ~~said~~ values accessible to the voice recognition system. Filing the values in a table has the advantage that an individual adjusting of various values of the input quantity to respectively different values of the system parameters can be carried out.

Another development consists in performing the setting during the operation of the voice recognition system. The advantage of this is that the adjustment of the voice recognition system can be individually adapted according to the respective requirements during the operation of the voice recognition system.

An additional development of the invention consists in specifying at least one of the following system parameters with the aid of the input quantity:

a) pruning threshold;
b) histogram pruning;
c) acoustic look-ahead
d) language model look-ahead;
e) threshold for selecting distance parameters that are to be computed.

Refer to the introduction of this specification for the meaning and function of these system parameters.

The cited system parameters represent a selection of possibilities. There are other imaginable system parameters, which can differ from those above depending on the respective voice recognition system.

It is also a development of the invention that the system parameters are weighted in consideration of their influence on a target quantity. The target quantity can be the accuracy of the voice recognition system or a speed of the voice recognition system (i.e. the speed at which the voice recognition process is carried out). In view of the respective target quantity, the system parameters can be weighted proportionally equally or differently in accordance with a predetermined weighting table.

In the context of another development, the input quantity is specified with the aid of an adjusting element.

The adjusting element advantageously comprises a one-dimensional degree of freedom with two limits, the first limit being converted as a maximum accuracy of the voice recognition system and the second limit being converted as a maximum speed of the voice recognition system.

It is also a development of the method that the adjusting element is represented on a computer as a slider and is operated with the aid of a keyboard, a touch-pad or a mouse.

The adjusting element can also be a rotary controller, a slider or a potentiometer.

In an additional development, the adjusting element is voice activated, the speech being evaluated by a voice recognition device, particularly the voice recognition system. The input quantity can be specified by voice input.

Another development of the method is that a fully automated specifying of the input quantity is carried out in the following steps:

A performance capability of the computer on which the voice recognition system is to run is determined using a program for performance assessment and is

stored as a performance index. In consideration of the performance index, the system parameters of the voice recognition system are automatically adjusted, thereby guaranteeing a high-performance voice recognition under real-time conditions, for example.

A program for determining the performance capability of the computer can be a prescribed loop, which runs through a definite number of iterations, the time for which iterations is measured. With the aid of a table, the performance index can be determined from the measured time. There are also programs that can be obtained commercially or as freeware which compute the performance of the computer and output a performance index as a rating of the performance capability.

A device for voice recognition is also set forth, which comprises a voice recognition system and which is realized with means for setting an accuracy of the voice recognition system, which means converts system parameters of the voice recognition system from an input quantity; that is, adjusts the voice recognition system and its numerous parameters with the aid of the input quantity.

It is advantageous here that such an adjusting of the system parameters with the aid of the input quantity can occur during the operation of the voice recognition system. This makes it possible for the user to easily adapt the number ~~plurality~~ of system parameters.

In one development the input quantity can be specified automatically. To this end, what is known as a performance index is computed with the aid of a device for measuring the performance of the computer on which the voice detection runs, and this index is used for adjusting the accuracy of the voice recognition system.

Another development provides that the input quantity can be prescribed by an adjusting element. A variety of adjusting elements (potentiometers, virtual control

units on the computer, and so on) whose setting directly determines the accuracy of the voice recognition system can be used for this.

Steps of the inventive method can expediently be carried out on the cited device.

Developments of the invention ~~emerge from the dependent claims.~~ A17

INS A18

Exemplifying embodiments of the invention are detailed below with the aid of the following Figures.

~~Shown are:~~

Figure 1 is a flow chart of a system architecture for a voice recognition system;

Figure 2 is a block diagram with steps of a method for setting the accuracy of a voice recognition system;

Figure 3 is a block diagram representing a linking of an input quantity to at least one system parameter via a mapping specification;

Figure 4 is a diagram representing different possible system parameters of the voice recognition system;

Figure 5 is a diagram ~~a sketch~~ showing the influence of the system parameters on a target quantity;

Figure 6 a diagram showing different possibilities for adjusting the input quantity;

Figure 7 a voice recognition system with a means for setting the accuracy.

INS A19

<u>Figure 1</u> shows a system architecture for voice recognition (voice recognition system).

The recognition of naturally spoken language is predicated on an appropriate formalism for representing knowledge. A complete voice recognition system encompasses several processing levels. These include acoustics-phonetics, intonation,

syntax, semantics, and pragmatics. The processing levels in the recognition are represented in Figure 1.

The natural speech signal 101 enters the voice recognition system. There, a feature extraction is performed in a component 102. After the feature extraction, sounds are recognized with the aid of known acoustic-phonetic units 103 (see block 104). This is a matter of computing acoustic distance parameters. After the sound recognition 104, the lexical decoding (word recognition) occurs in block 106 with the aid of the pronunciation model, or respectively, word lexicon 105, and then a syntax analysis 108 with the aid of the language model 107 that encompasses the grammar. The word recognition 106 and the syntax analysis 108 represent the search for a correspondence for the speech signal. Finally, in a block 110 a semantic post-processing is performed, whereby contextual knowledge and pragmatics 109 are taken into account, and this finally results in the speech that is recognized by the voice recognition system.

Figure 2 shows a block diagram representing the steps of a method for setting the accuracy of a voice recognition system.

In a step 201, the accuracy of the voice recognition system is specified by a prescribable input quantity. Next, in a step 202 values for system parameters of the voice recognition systems are computed using these input quantities. Lastly, in a step 203 the voice recognition system is adjusted using the computed values.

Figure 3 shows a block diagram representing a linking of an input quantity to at least one system parameter via a mapping specification.

The cited input quantity 301 is mapped onto the system parameters SP 303 of the voice recognition system with the aid of a mapping specification 302. One input quantity 301 is advantageously allocated to several system parameters via the mapping specification. On the basis of this mapping specification 302, the voice

recognition system is adapted by specifying an input quantity; i.e., several system parameters SP are influenced by modifying one input quantity 301. The mapping specification 302 advantageously has the form of a table in which a column contains possible input quantities 301, and in a row of these columns several values for system parameters SP303 are allocated to the respective input quantity. The mapping specification 302 consists in the search for the entry (row) in the table that is allocated to the input quantity 301 and in the transfer of the found values for system parameters SP303 to the voice recognition system.

Figure 4 shows a diagram representing different possible system parameters of the voice recognition system. The system parameters SP of the voice recognition system, represented in a block 401, include at least one of the following parameters:

a) pruning threshold 402
b) histogram pruning 403
c) acoustic look-ahead 404
d) the language model look-~~aheadm [sic]~~ ahead 405
e) threshold for distance parameter 406.

Overall it is possible to imagine additional system parameters of the voice recognition system for performing a setting via the input quantity 301, indicated by the block 407.

Figure 5 shows a sketch representing the influence of the system parameters on a target quantity.

The system parameters SP (see block 501) exert influence on a target quantity ZG (see block 502). As described above, there are several system parameters SP1, SP2, and so on, which are represented in a block 503. Here, each individual system parameter SPi (i=1,2,...) influences the target quantity ZG by means of a weight Gi that is provided

for the respective system parameter SPi (represented in a block 504). By weighting the system parameters SP, it is possible to exert a variable influence on the target quantity ZG, depending on the system parameter SP.

Figure 6 shows a diagram representing the different possibilities for setting the input quantity. The setting of the input quantity, represented in the block 601, is accomplished using operating components of the computer R (see block 602), using control components 603, or using the computer itself (see block 604). To this end, the computer R comprises at least one means for setting the input quantity, such as a keyboard 605, a mouse 606, a touch-pad 607, or voice input 608 via the voice recognition system. Possible components for control, represented in the block 603, include a rotary control 609, a slider 610, or some other control 611, preferably a potentiometer. In addition, the computing power of the computer is automatically determined with the aid of a program that runs on the computer, and the input quantity for setting the voice recognition system is specified accordingly. It is thus guaranteed that an automatically adjusted computer guarantees a quality in the voice recognition process that is commensurate with its computing power. A compromise is automatically found between high quality in the voice recognition at the expense of computing power, or respectively, real-time recognition of the language, and rapid voice recognition with a correspondingly low demand for computing power but with significant compromises of the quality of the voice recognition.

Figure 7 shows a device consisting of a voice recognition system 701 and a means for adjusting the accuracy of the voice recognition system 702.

System parameters of the voice recognition system are specified by an input quantity using the means for adjusting the accuracy 702. To this end, the input quantity is allocated to a plurality of system parameters (pruning threshold, histogram pruning,

acoustic look-ahead, language model look-ahead, threshold for selecting distance parameters to be calculated, etc.), preferably with the aid of a prescribed table.

The input quantity can alternatively be specified using an adjusting element 703 or a device for determining the performance of the computer 704. Refer here to Figure 6 and the possibilities cited there for setting the input quantity.

INS A20

The folowing publications were cited in this document:

[1] A. Hauenstein: "Optimierung von Algorithmen und Entwurf eines Prozessors für die automatische Spracherkennung"; Lehrstuhl für Integrierte Schaltungen, Technische Univerität München, Dissertation, 7-19-93, Chapter 2, pp.13-26.

[2] A. Hauenstein: "Optimierung von Algorithmen und Entwurf eines Prozessors für die automatische Spracherkennung"; Lehrstuhl für Integrierte Schaltungen, Technische Univerität München, Dissertation, 7-19-93, Chapter 3.3.3; pp. 40-43

[3] Volker Steinbiss, Bach-Hiep Tran, Hermann Ney: "Improvements in Beam Search", Proc. Intl. Conf. Speech and Language Processing; Yokahama 1994, pp. 2143-2146.

[4] M. Niemöller, A. Hauenstein, E. Marschall, P. Witschel, U. Harke: "A PC-based Real-Time Large Vocabulary Continuous Speech Recognizer for German", Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing; Munich 1997.

[5] A. Hauenstein: "Optimierung von Algorithmen und Entwurf eines Prozessors für die automatische Spracherkennung"; Lehrstuhl für Integrierte Schaltungen, Technische Univerität München, Dissertation, 7-19-93, Chapter 3.5.1; pp. 65-69

[6] S. Ortmanns, A. Eiden, H. Ney, N. Coenen: "Look-Ahead Techniques for Fast Beam Search"; Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing; Munich 1997, pp. 1783-1786

[7] E. Bocchieri: "Vector Quantization for the Efficient Computation of Continuous Density Likelihoods", Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing; 1993:II-692 to II-695